Electronic Version
Stylesheet Version v1.1.1

SYSTEM AND METHOD FOR COMPUTING A MEASURE OF SIMILARITY BETWEEN DOCUMENTS

Abstract

A measure of similarity between two documents is computed. In computing the measure of similarity, a first list of rated keywords extracted from the first document and a second list of rated keywords extracted from the second document are received. The first and second lists of keywords are used to determine whether the first document forms part of the second document using a first computed percentage indicating what percentage of keyword ratings in the first list also exist in the second list. A second percentage is computed that indicates what percentage of keyword ratings along with a set of their neighboring keyword ratings in the first list that also exist in the second list when the first percentage indicates that the first document is included in the second document. The first percentage is used to specify the measure of similarity when the second percentage is greater than the first percentage.